



U.S. Department of Energy Office of Science

Environmental Remediation Sciences Program

Community Proteogenomics:

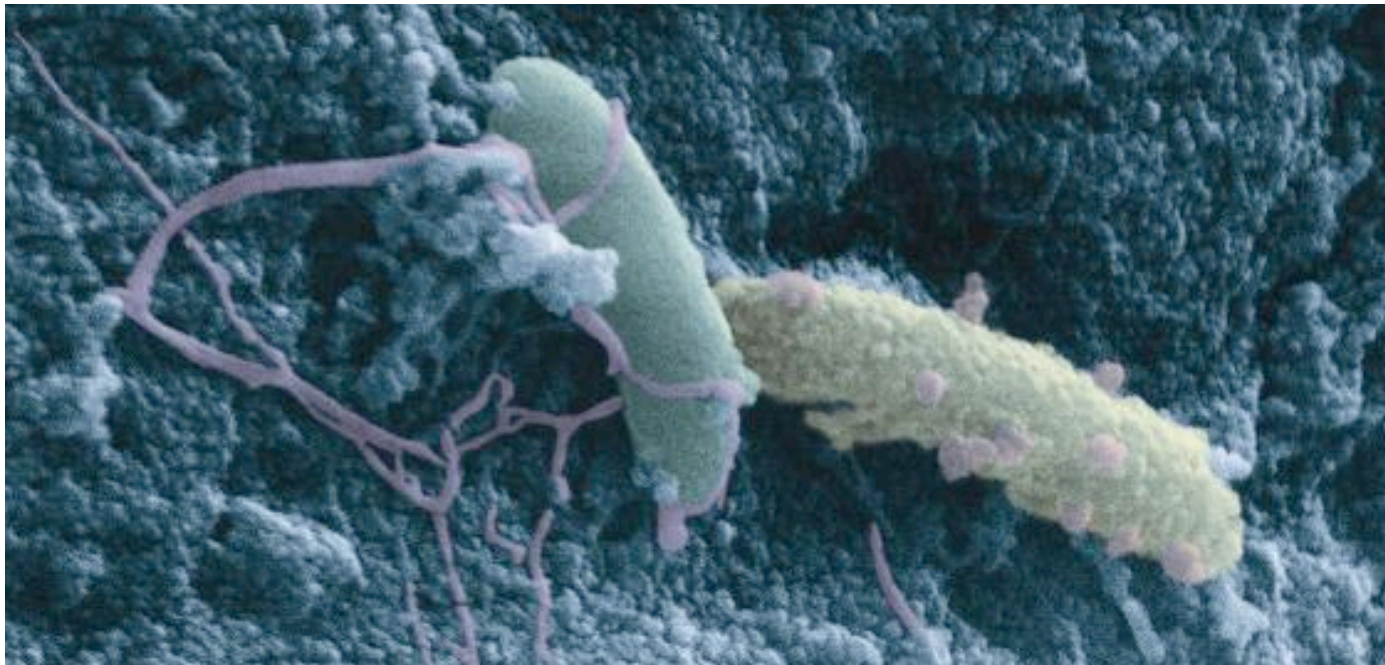
background and application to the Rifle
Bioremediation project

Rifle project proteogenomics subgroup
Jill Banfield, presenter

Goal:

to understand the functioning of natural microbial consortia

- study of coexisting organisms, not monocultures
- cultivation independent



Molecular foundation:

Genomes recovered from environmental samples

Metagenomics - data of data genomics or transcendental genomics

Environmental genomics - the study of environmentally-derived DNA

Community genomics - genomic studies of microbial communities

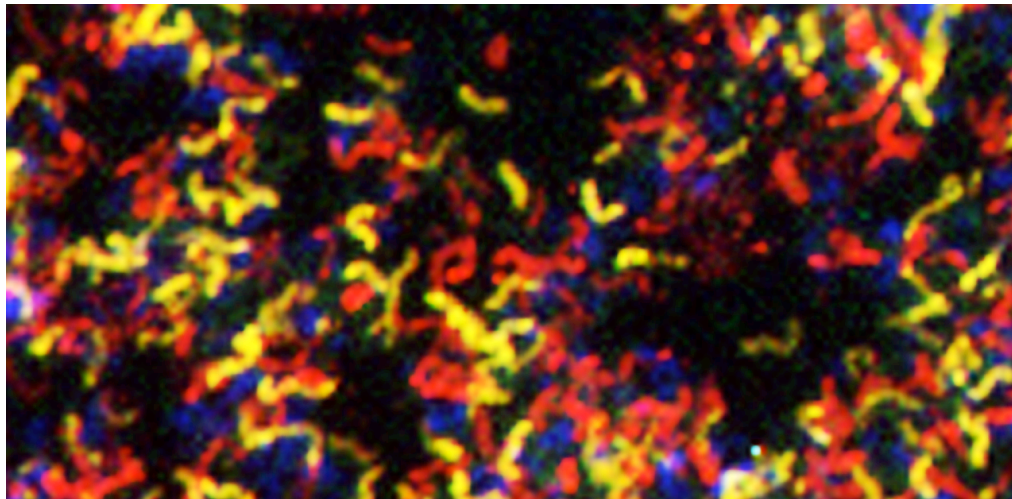


Image: V. Deneff, UC Berkeley

Community genomic analysis

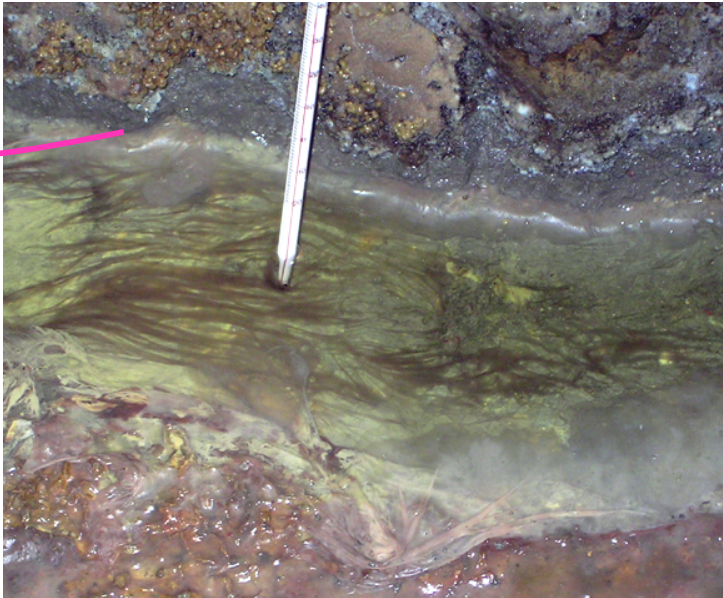
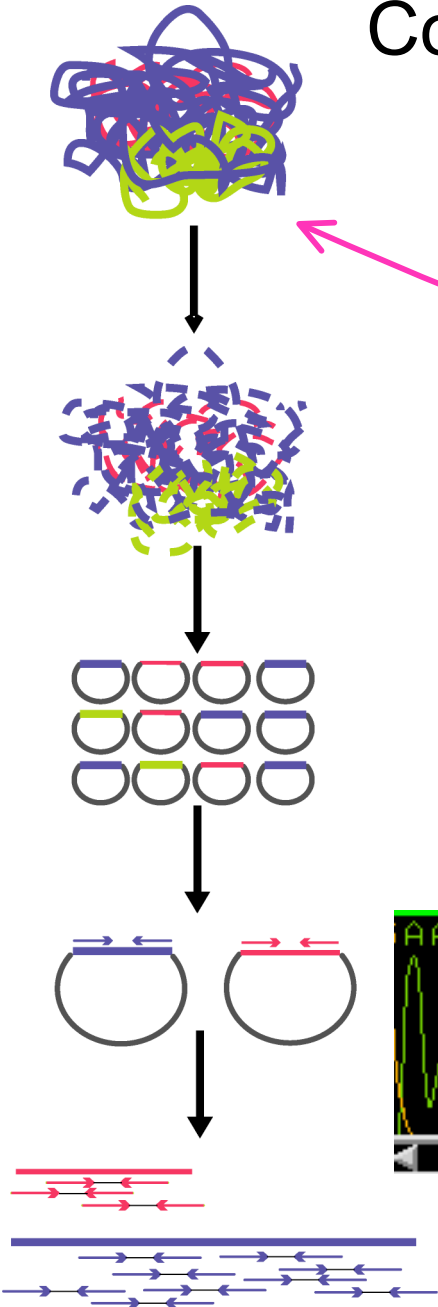
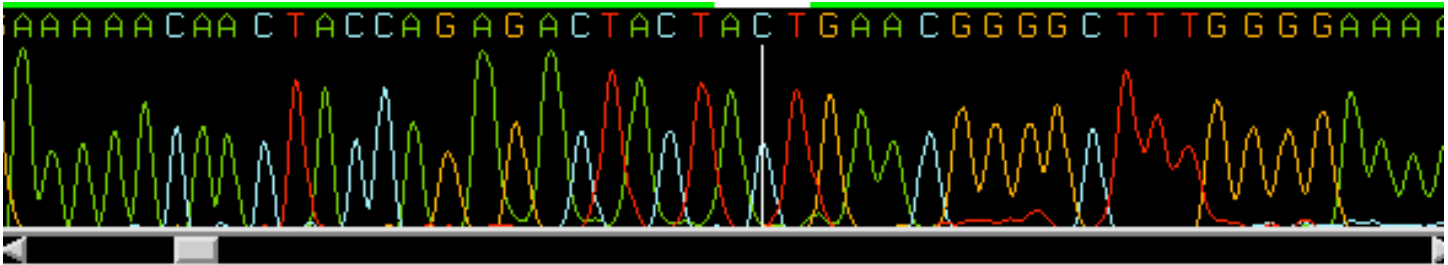
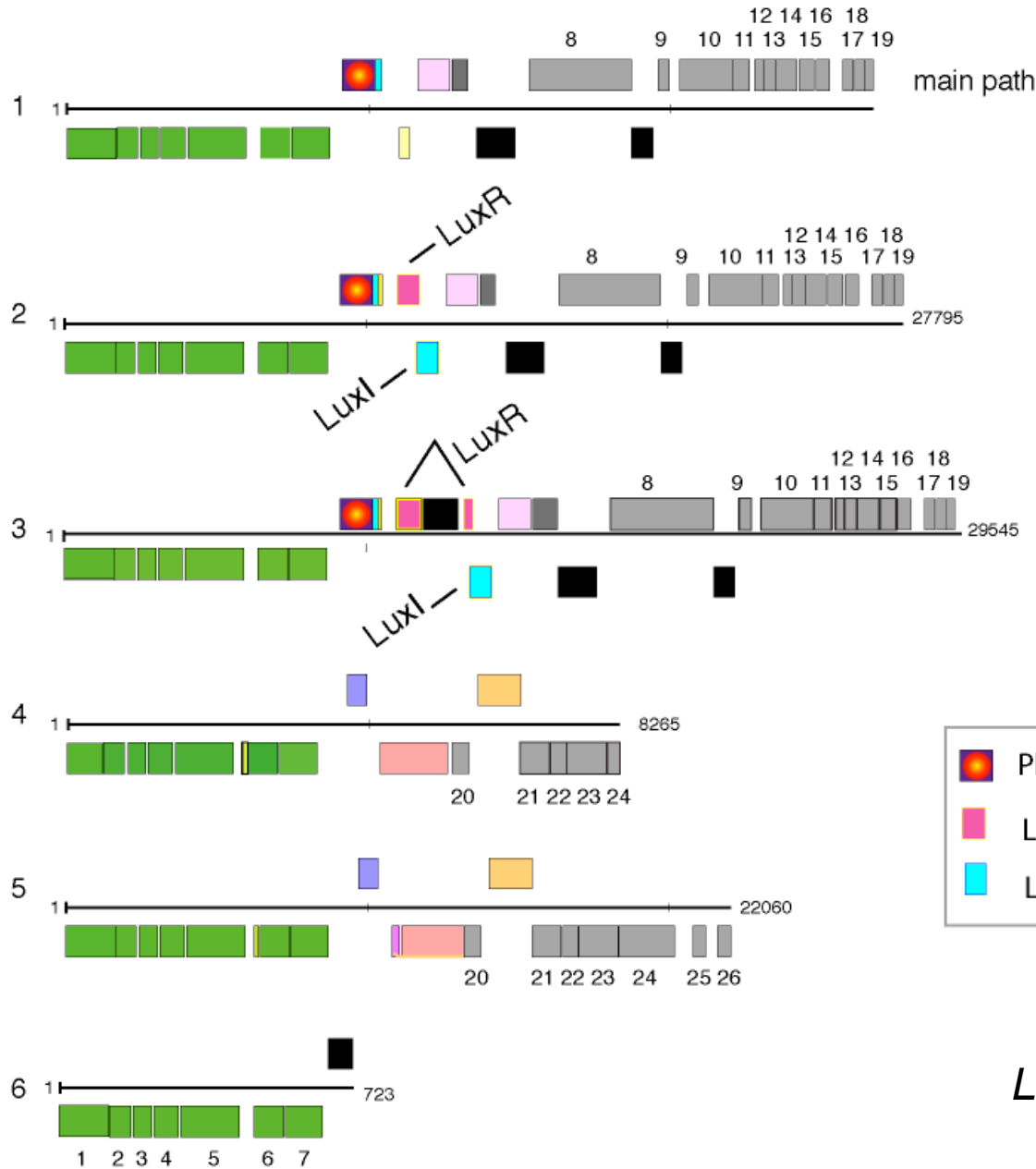


Photo: Rudy Carver

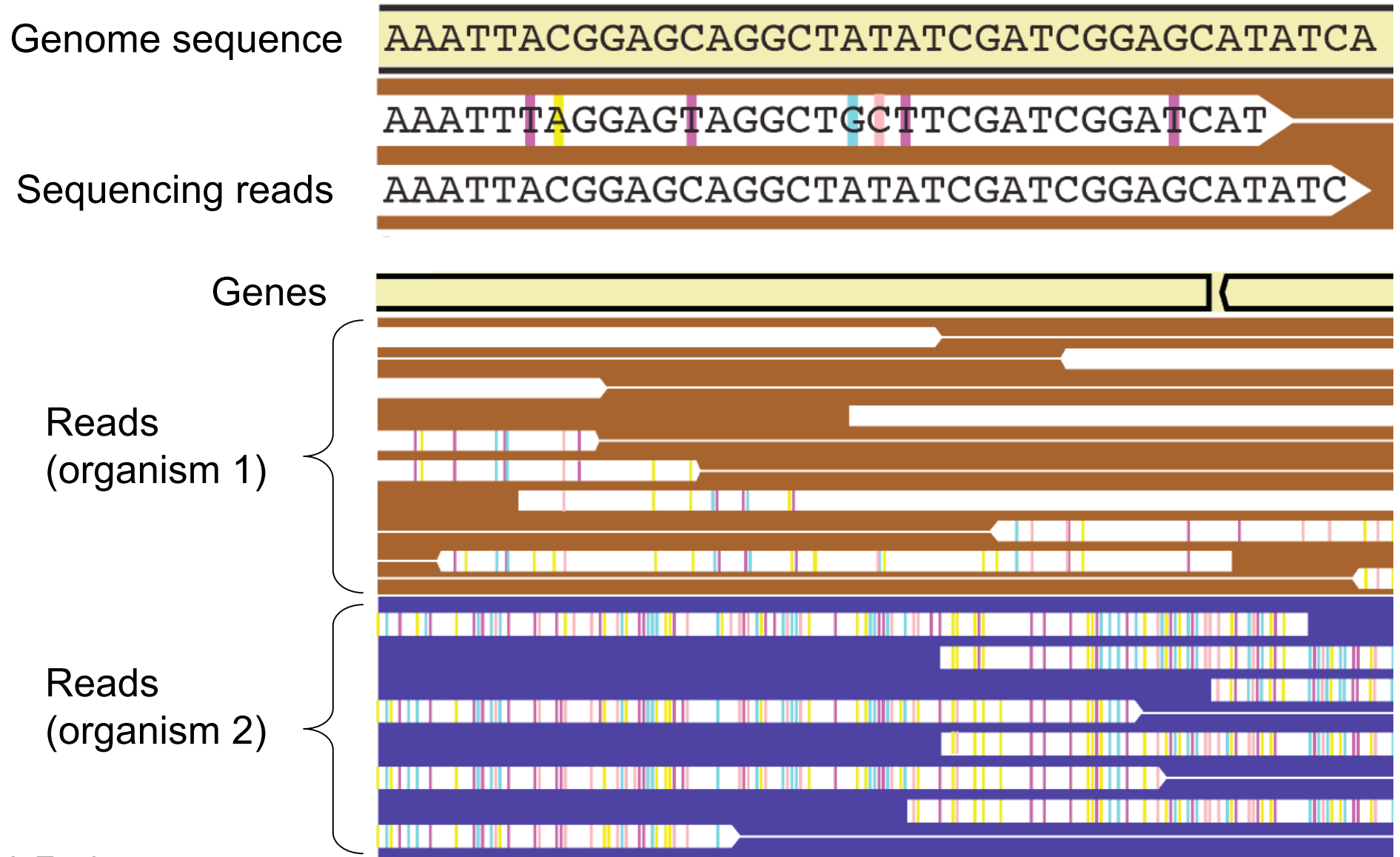


STRAIN DIFFERENCES IN GENE CONTENT

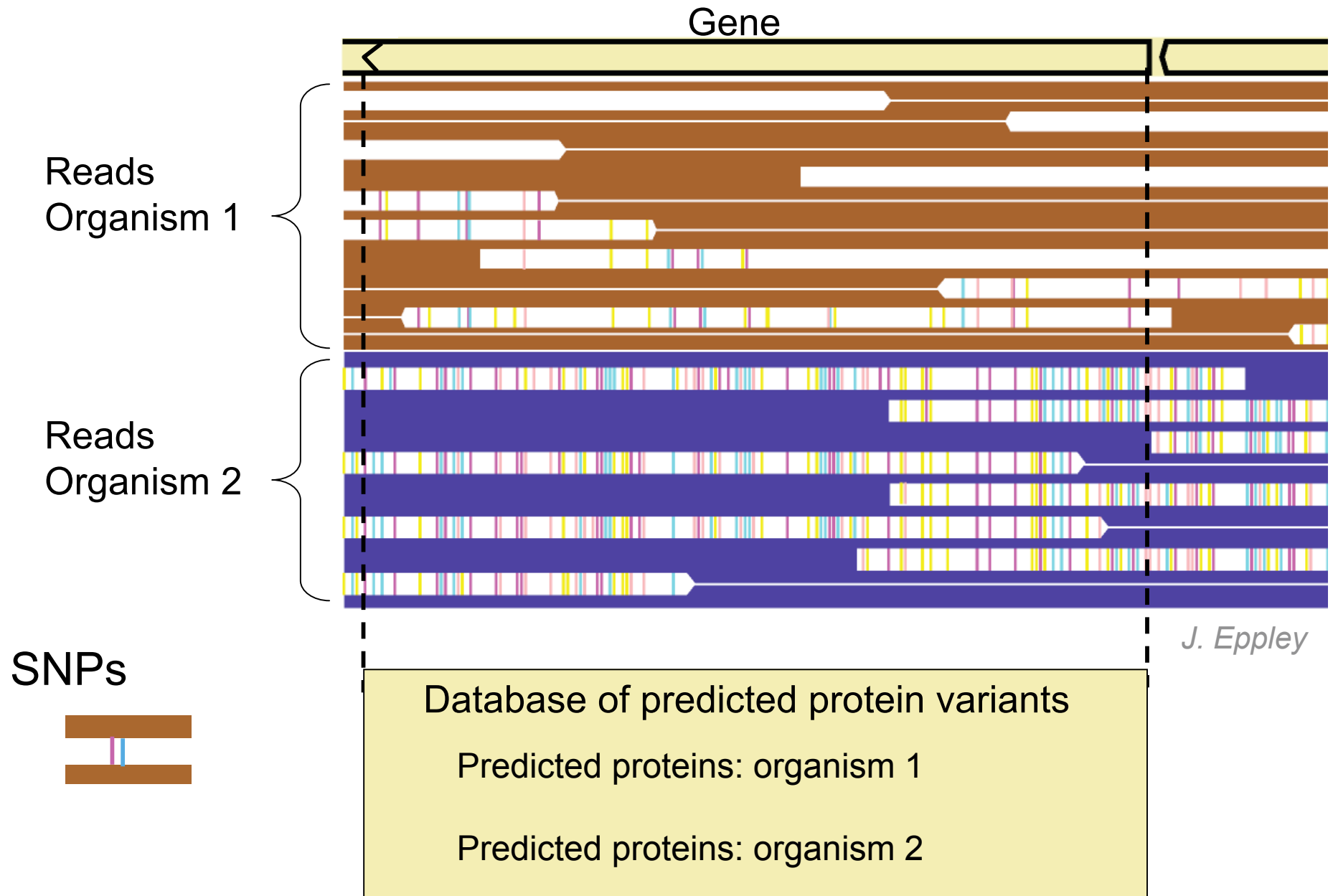


Lux genes only in minor *Leptospirillum* group II variants

Definition of populations: variation within vs. between clusters:



Reconstruction of protein variants using “Strainer” (Eppley et al.)

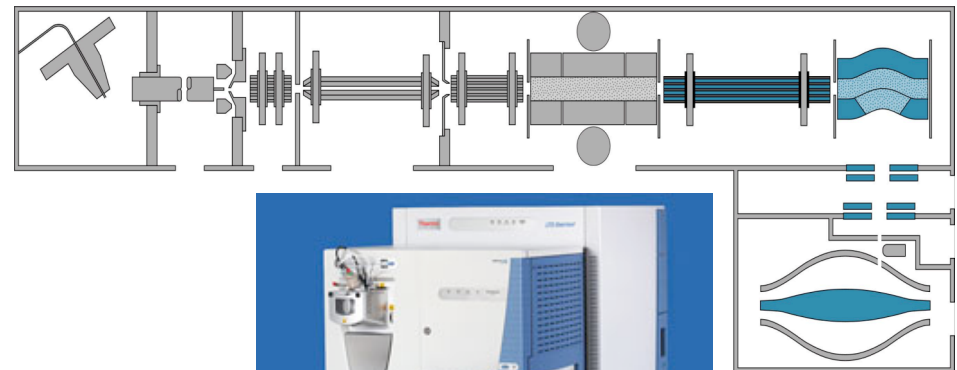


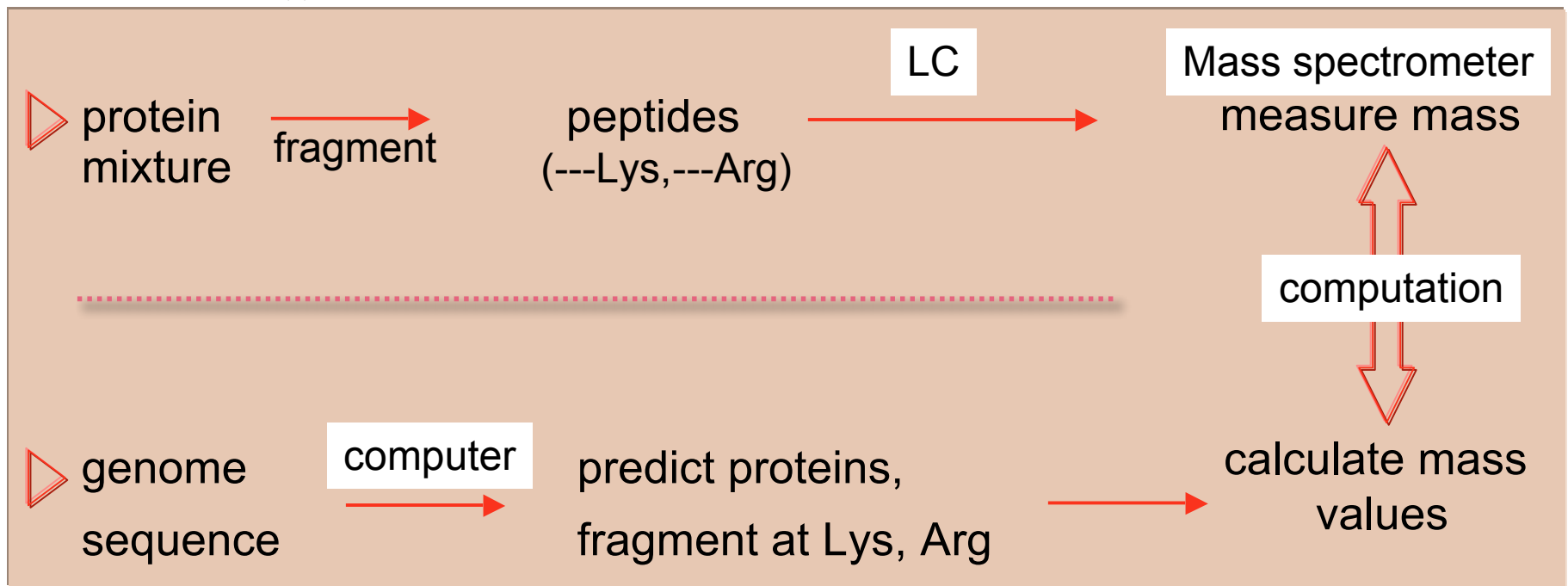
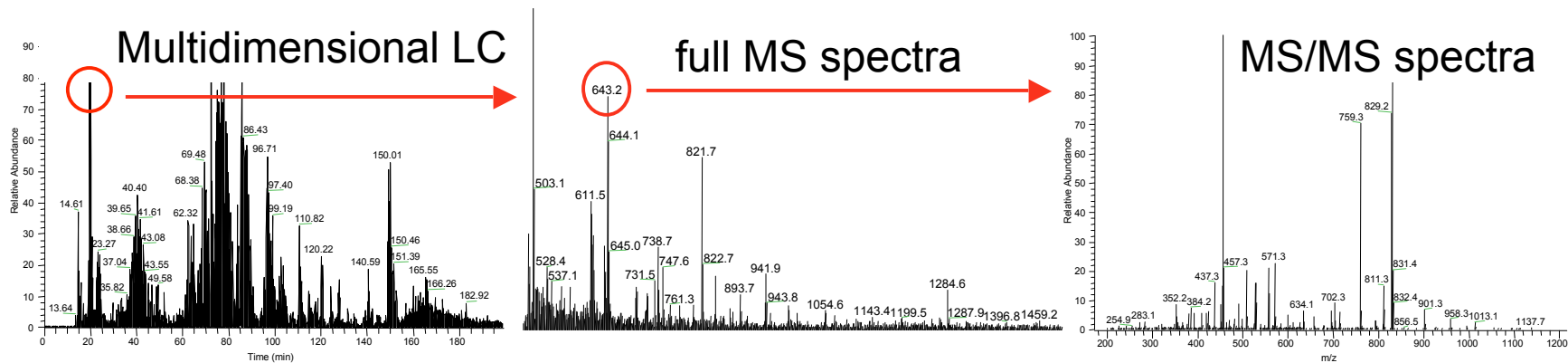
Global proteomics: Simultaneous characterization of all cell proteins

Peptide identification based on:

MS/MS fragmentation patterns of chromatographically-separated peptides; identification usually by comparison with protein fragmentation patterns predicted from genomic data

- In some cases, coupled to very high measured mass accuracy measurements of peptides in elution profiles





Leptospirillum group II_scaffold_14_GENE_20

2032 proteins
from one biofilm

MNKWAGAVLGTVTLLGLLSATAYSAELDILKPNRVPADQIAAAKAMKPPFPVTA
AVIAKGKEVFNGAGTCYTCHGVGGK**GDGPGAAGMDPS**PRFTNHQFDQVRTAGE
MVVVVSNGLSPLQPMVGFVSAGITDKQAWAVMYERSLGC GGDMDC.....

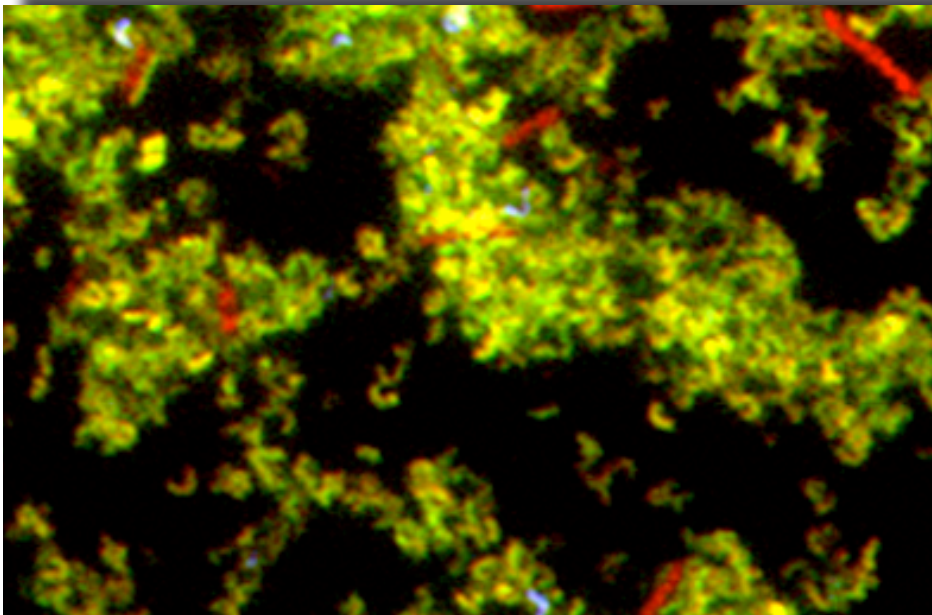
LOW DIVERSITY COMMUNITIES AS MODEL SYSTEMS



Most biofilms dominated by:

Leptospirillum group II
Leptospirillum group III

Ferroplasma types I and II
A, B, C, D, E, G, H, I-plasma
ARMAN 1,2,3,4 at low
abundance



Blue = NEAR-COMPLETE
COMPOSITE GENOMES
RECONSTRUCTED
OR ONGOING

Tyson et al. *Nature* (2004)
Allen et al. *PNAS* (2007)
Lo et al. *Nature* (2007)

SOME SIMPLE METRICS

For the most abundant organism - single sample:

~ 52% predicted proteins detected in a single sample

For the most abundant organism - across 6 samples:

68% of all predicted proteins

80% of core chromosomal proteins

34% of integrated plasmid/phage proteins

Overall, proteins identified across multiple samples:

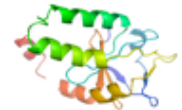
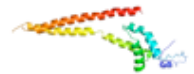
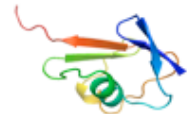
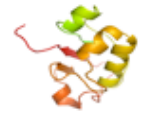
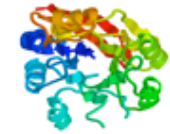
68% *Leptospirillum* group II

41% *Leptospirillum* group III

37% Gplasma

20% *Ferroplasma* type I

20% *Ferroplasma* type II



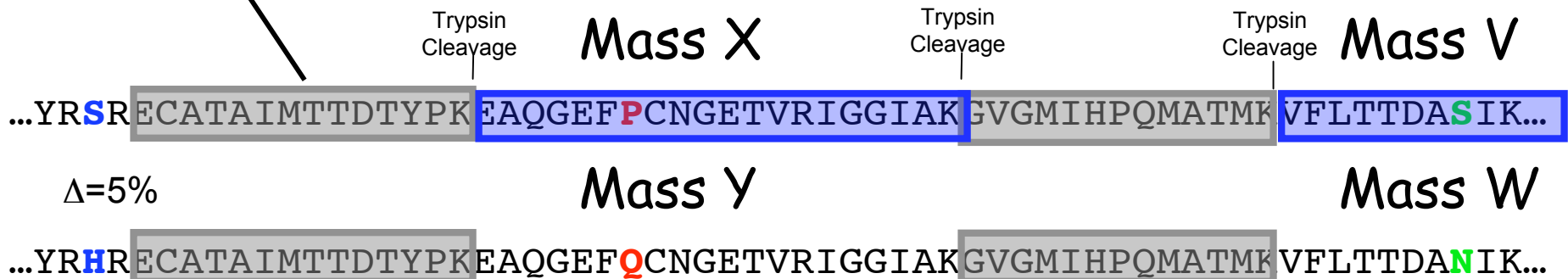
Biofilms: 2,275 - 3,989 proteins with ~1% false positive identification rate)

STRAIN-RESOLVED PROTEOMICS

Which protein variant
- strain 1 or strain 2?

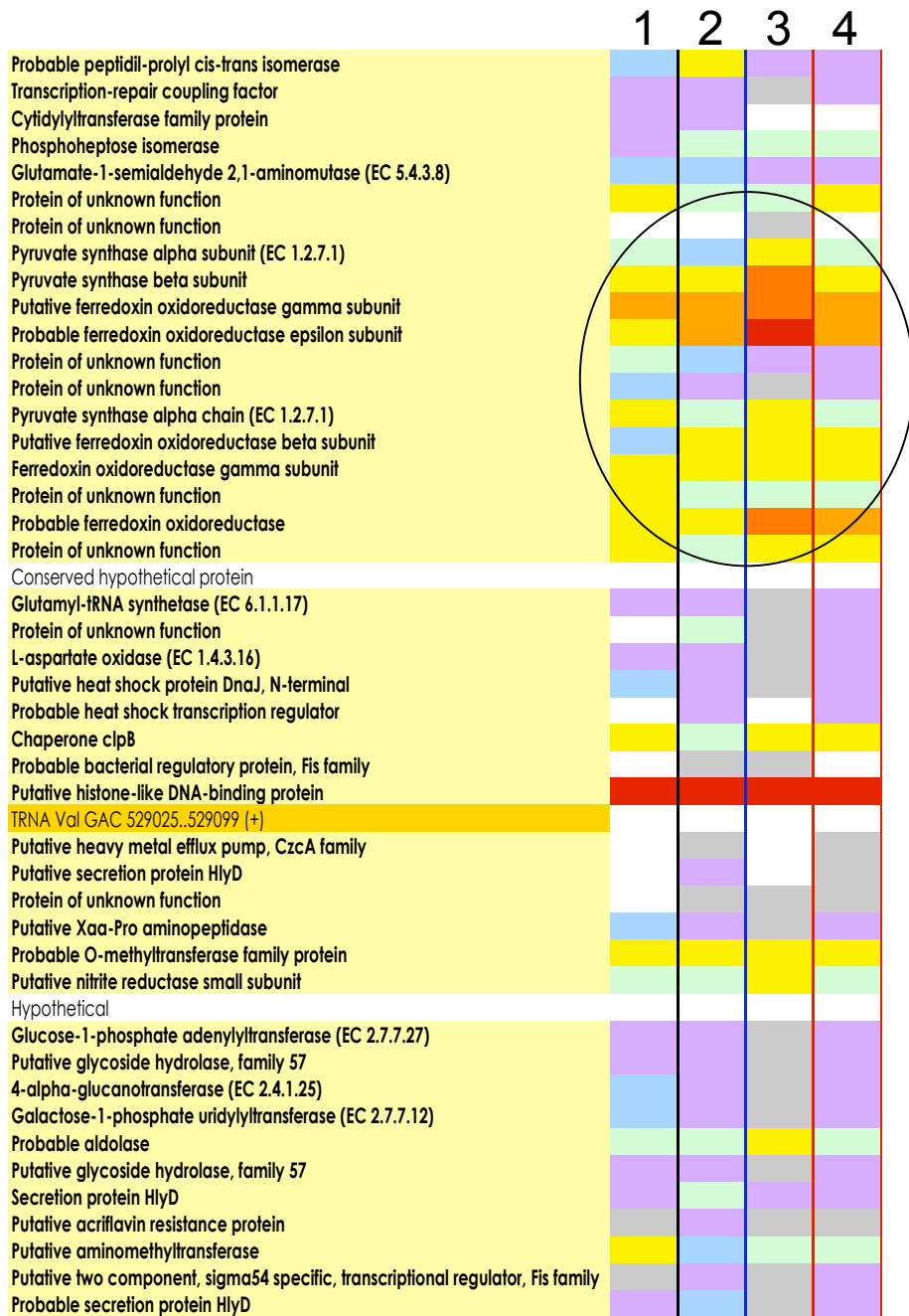


Sequence of protein
from strain 1



Sequence of protein
from strain 2

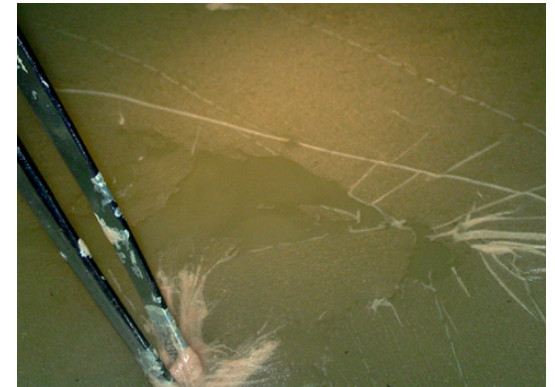
Proteome changes over space and time:



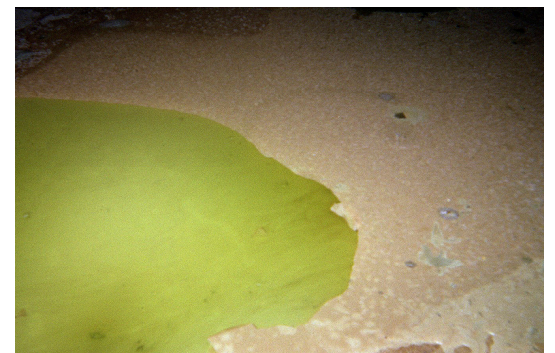
1 Bioreactor
pH 0.95, 37°C



2 ABend
pH 1.07, 43°C



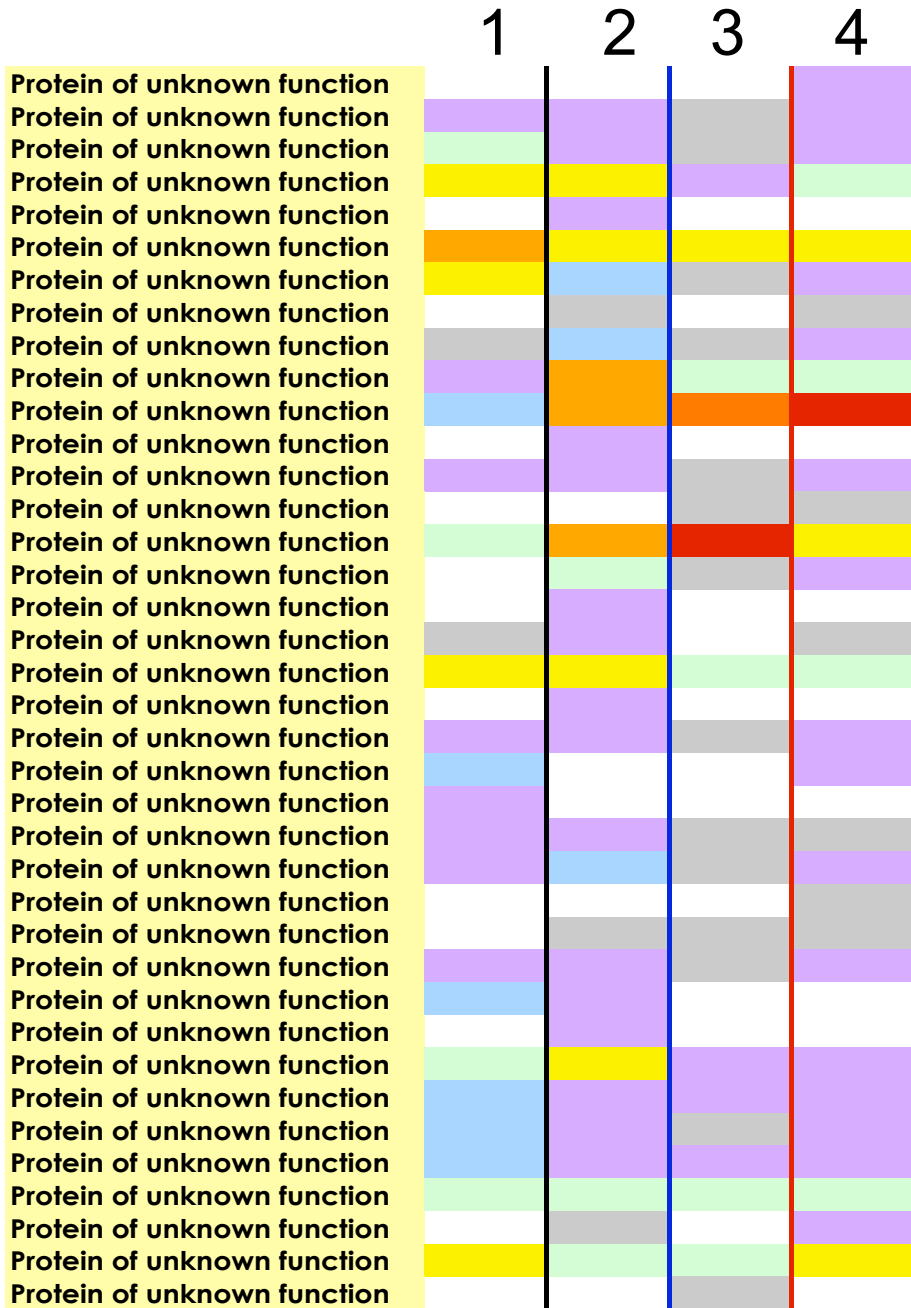
3 ABfront
pH 0.99, 39°C



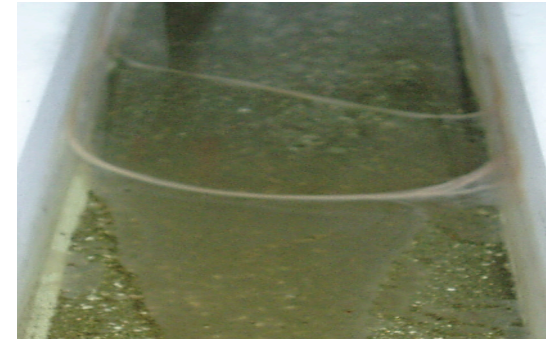
4 UBA
pH 1.28, 38°C



Proteins of unknown function



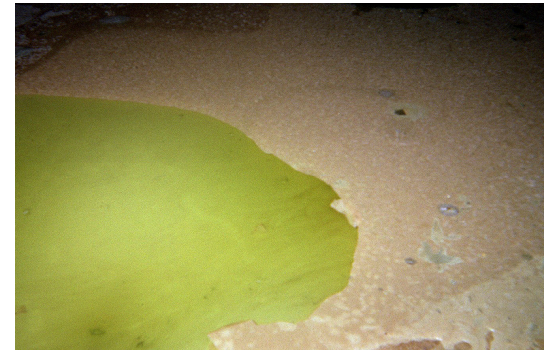
1 Bioreactor
pH 0.95, 37°C



2 ABend
pH 1.07, 43°C

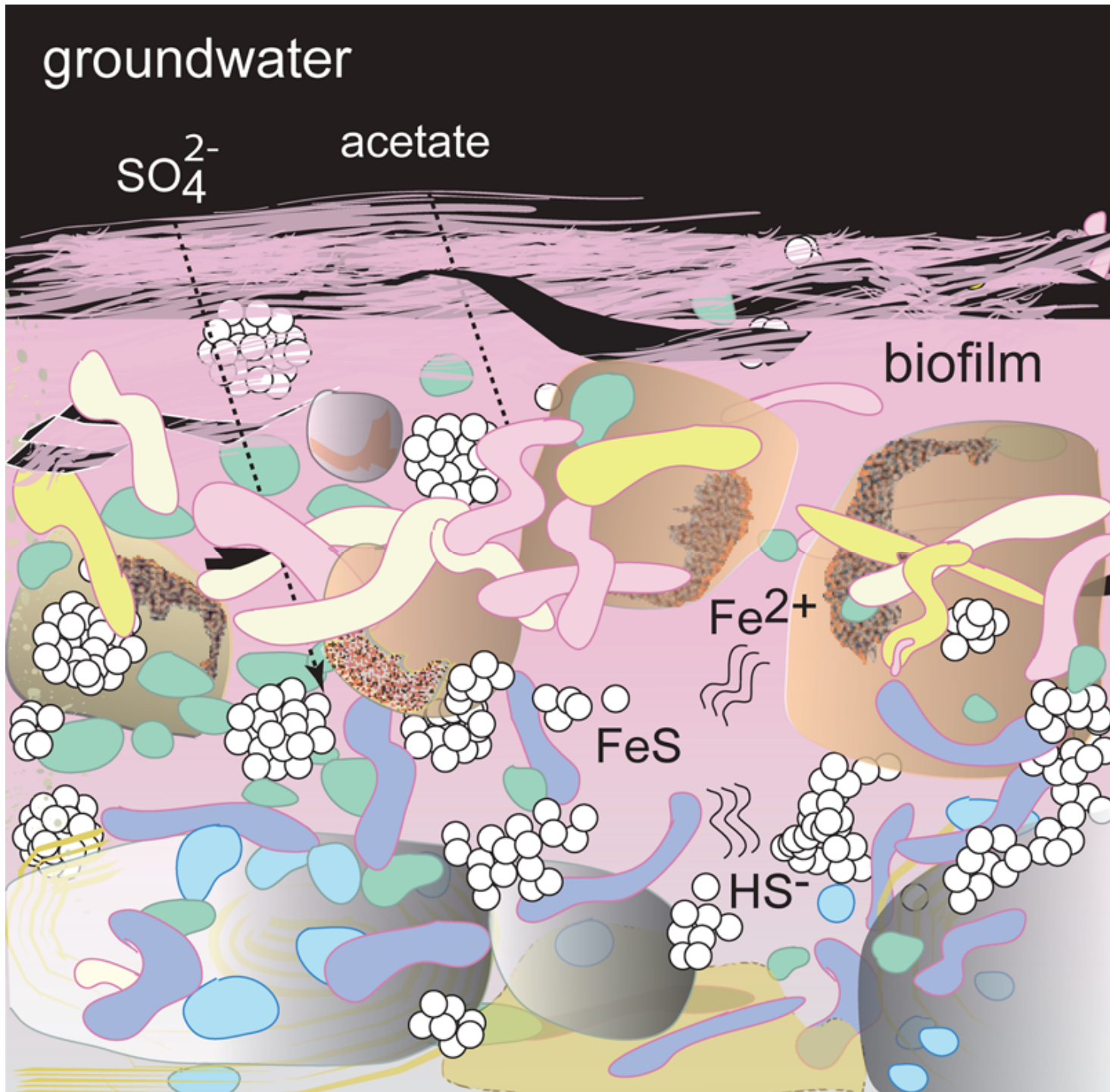


3 ABfront
pH 0.99, 39°C

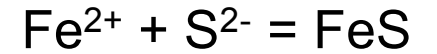


4 UBA
pH 1.28, 38°C





Microbial
reduction of
ferric iron-
bearing
minerals



Microbial
sulfate
reduction

Old Rifle Aquifer Sediment Columns Studies

Acetate: 3 mM, Sulfate: 9 mM, Flow Rate: 0.3 m/day



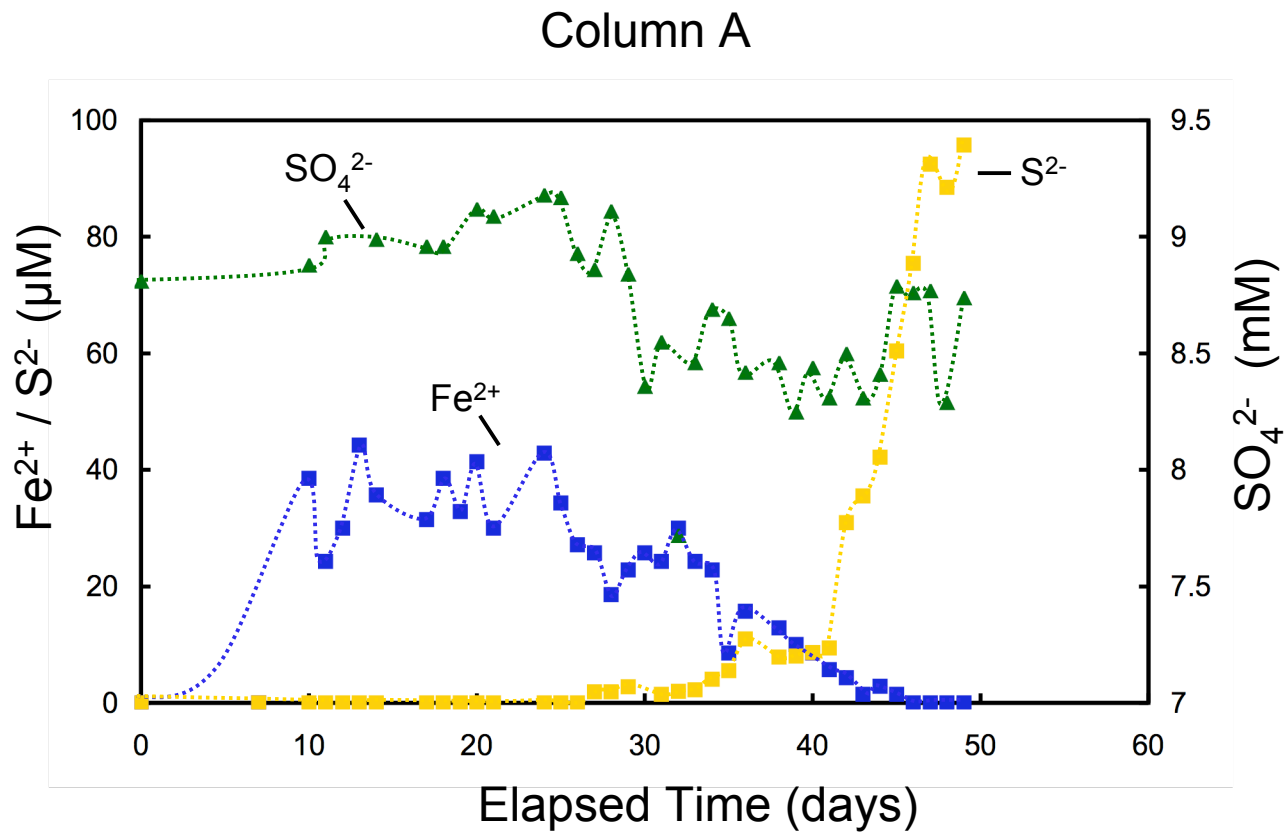
Start

48 days

Mike Wilkins, Ken Williams

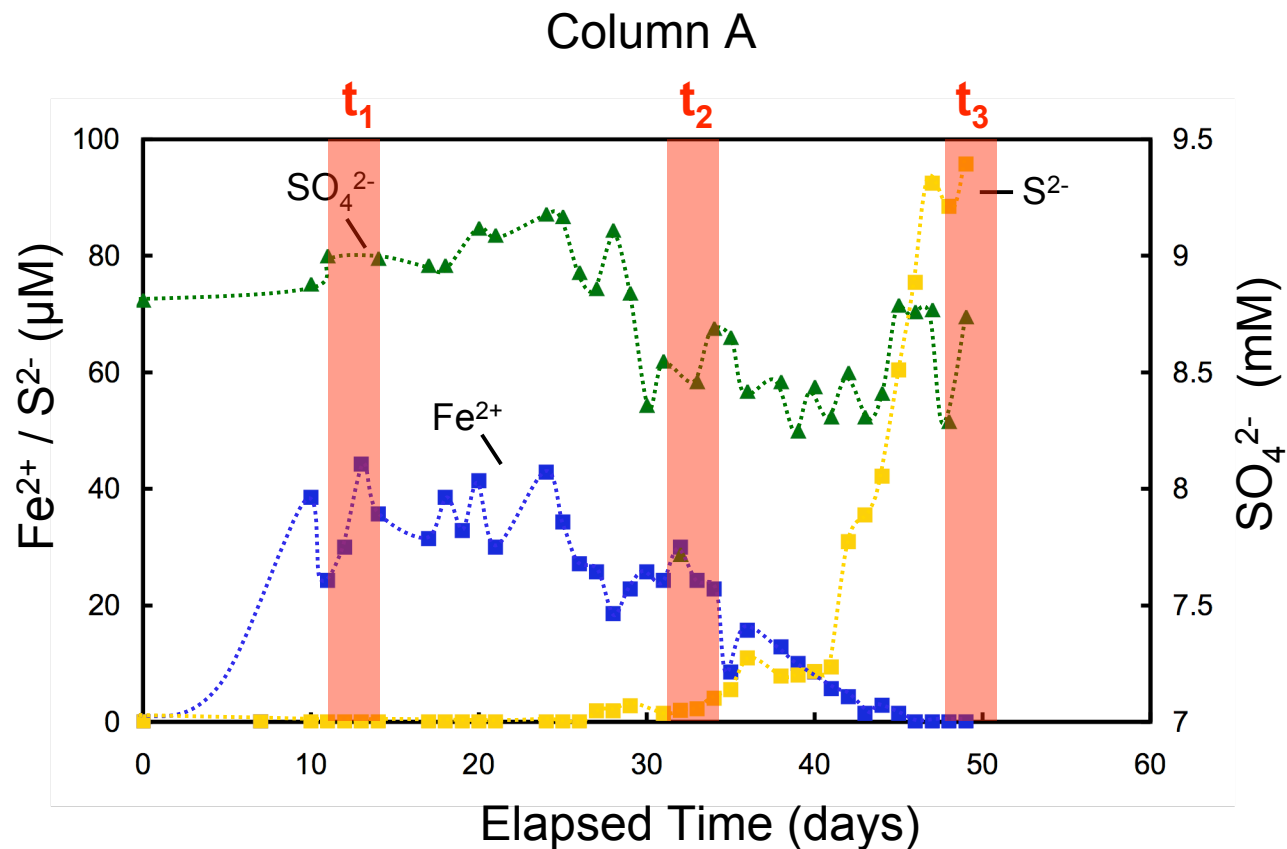
Column Studies

Protein sampling intervals *and locations* determined by pre-determined geochemical - *and geophysical* - indicators



Column Studies

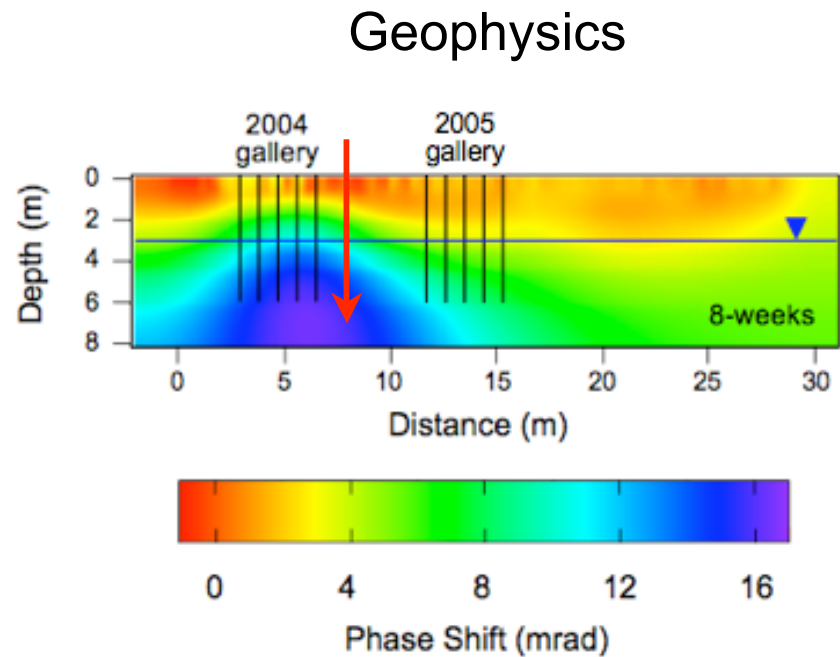
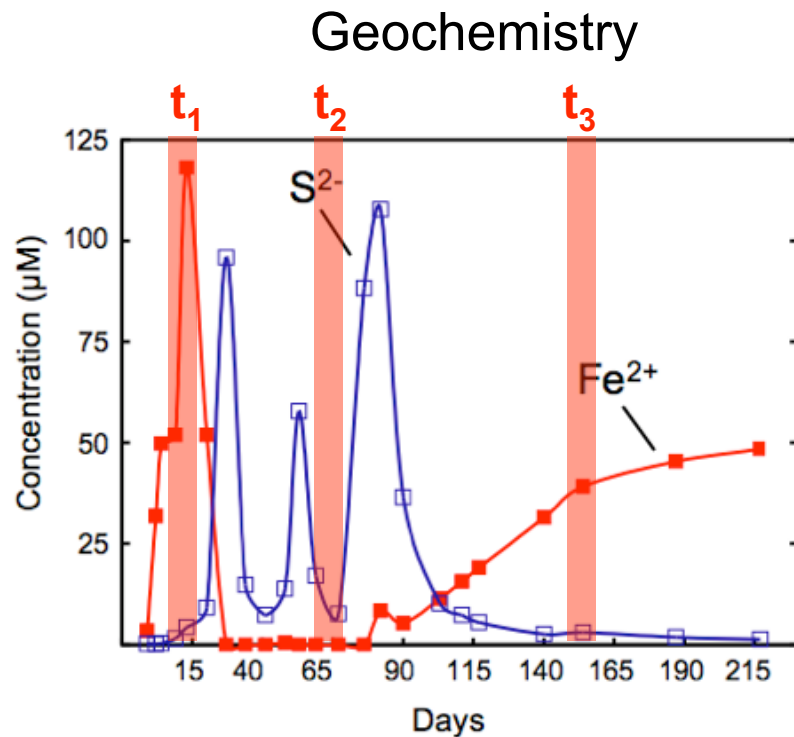
Protein sampling intervals **and locations** determined by pre-determined geochemical - and *geophysical* - indicators



Field Studies

Protein sampling intervals **and locations** determined by pre-determined geochemical - and *geophysical* - indicators

2006 Old Rifle Field Data



Rifle sediment proteomics

Proteins to be recovered from batch and column sediments, tubing, filters



Batch Incubations



Tubing / Effluent

M. Wilkins and K. Williams

Protein extraction: protocols for sediment under development

Metagenomic data: from Holmes and Lovley and sequencing to be requested from JGI

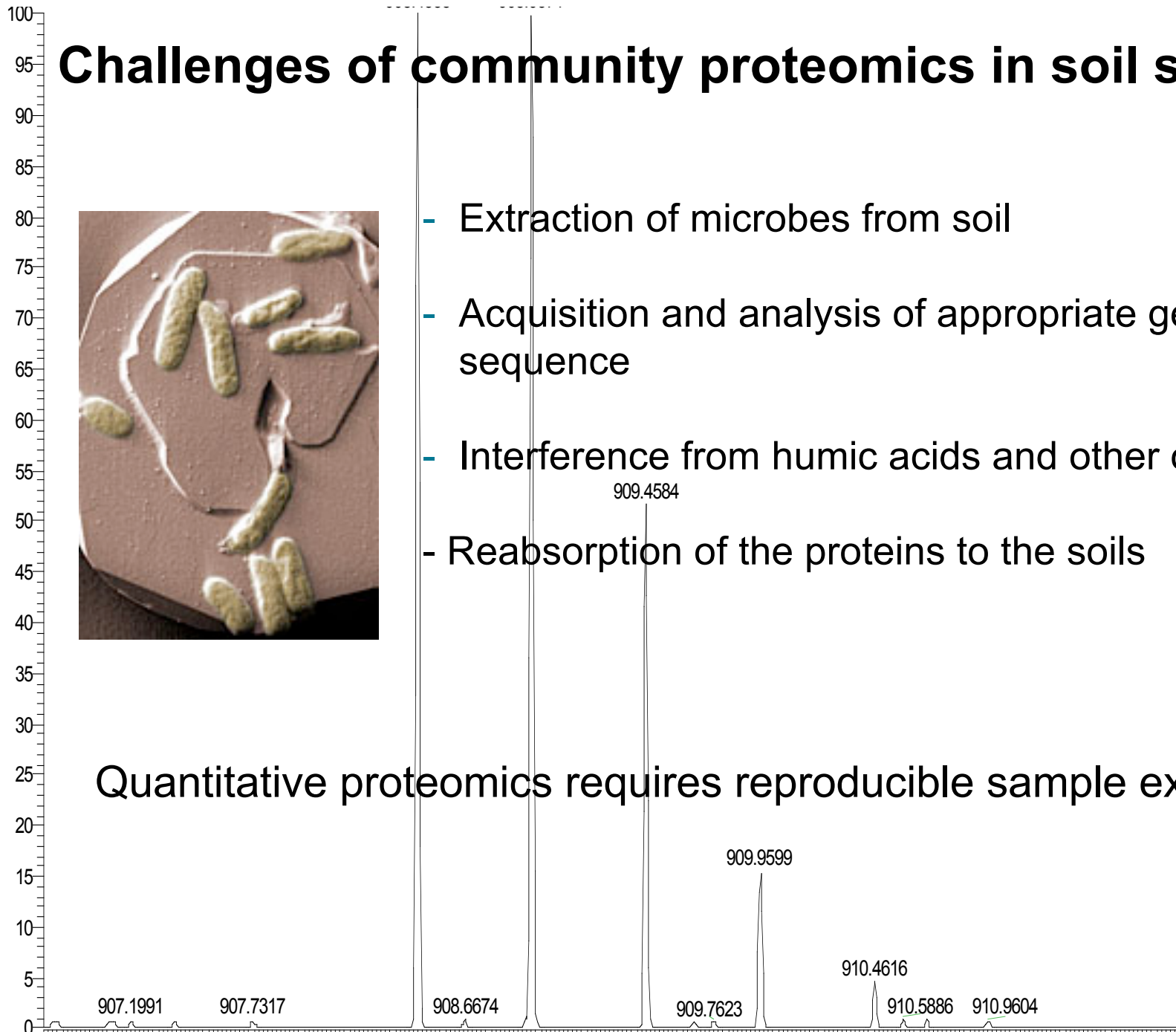
Test case: filtered groundwater samples from Rifle, summer 2007

Challenges of community proteomics in soil samples



- Extraction of microbes from soil
- Acquisition and analysis of appropriate genomic sequence
- Interference from humic acids and other organics
- Reabsorption of the proteins to the soils

Quantitative proteomics requires reproducible sample extraction





U.S. Department of Energy Office of Science

Environmental Remediation Sciences Program

Rifle project proteogenomics subgroup:

Mike Wilkins, UC Berkeley

Ken Williams, UC Berkeley

Greg Dick, UC Berkeley

Nathan VerBerkmoes, ORNL

Mary Lipton, PNNL

Dawn Holmes, U Mass

Robert Hettich, ORNL

Derek Lovely, U Mass

Phil Long, PNNL

Jill Banfield, UC Berkeley



AMD proteogenomics research:

Nathan VerBerkmoes, Vincent Deneff, Michael Thelen, Robert Hettich, Jill Banfield and collaborators